Applied Mathematical Sciences, Vol. 7, 2013, no. 13, 609 - 614

The Swap Common Superstring Problem

Anna Gorbenko

Department of Intelligent Systems and Robotics
Ural Federal University
620083 Ekaterinburg, Russia
gorbenko.ann@gmail.com

Vladimir Popov

Department of Intelligent Systems and Robotics
Ural Federal University
620083 Ekaterinburg, Russia
Vladimir.Popov@usu.ru

Abstract

In this paper we consider an approach to solve the swap common superstring problem. This approach is based on an explicit reduction from the problem to the satisfiability problem.

Keywords: swap common superstring, **NP**-complete, satisfiability

The algorithmic aspects of different problems of finding regularities are thoroughly studied in theoretical computer science (see e.g. [1] – [15]). In particular, the swap common superstring problem was proposed in [16].

Let

$$\Sigma = \{a_1, \ldots, a_m\}$$

be a finite alphabet. Let

$$\mathcal{S} = \{S_1, \ldots, S_n\}$$

be a collection of strings over Σ. For simplicity, we use $S[i]$ to denote the ith letter in string S, and $S[i, j]$ to denote the substring of S consisting of the ith letter through the jth letter. The length of a string S is the number of letters in it and is denoted as $|S|$. Let

$$\#occ(X, Y) = |\{i \mid X = Y[i, j]\}|.$$

The decision version of the swap common superstring problem can be formulated as following.

THE SWAP COMMON SUPERSTRING PROBLEM (SWCS):

INSTANCE: *A collection $\mathcal{S}$ of strings over Σ, a string T, and a positive integer k.*

QUESTION: *Is there a string S such that*

$$|S| = |T|,$$

$$\{S[i], S[i+1]\} = \{T[i], T[i+1]\},$$

for all $1 \leq i < |S|$, *and*

$$|\{i \mid \#occ(S_i, S) \geq 1\}| \geq k?$$

The problem SWCS is **NP**-hard [16]. Encoding different hard problems as instances of SAT and solving them with efficient satisfiability algorithms has caused considerable interest (see e.g. [17] – [36]). In this paper, we consider an approach to solve the SWCS problem. Our approach is based on an explicit reduction from the problem to the satisfiability problem.

Let

$$\varphi[1] = \wedge_{1 \leq i \leq |T|} \vee_{1 \leq j \leq m} x[i, j],$$

$$\varphi[2] = \wedge_{\substack{1 \leq i \leq |T|, \\ 1 \leq j[1] < j[2] \leq m}} (\neg x[i, j[1]] \vee \neg x[i, j[2]]),$$

$$\varphi[3] = \wedge_{\substack{1 \leq i \leq |T|-1, \\ 1 \leq j \leq m, T[i] \neq a_j, \\ 1 \leq s \leq m, T[i+1] \neq a_s}} ((\neg x[i, j] \vee \neg x[i+1, s]) \wedge (\neg x[i, s] \vee \neg x[i+1, j])),$$

$$\varphi[4] = \wedge_{1 \leq i \leq k} \vee_{1 \leq j \leq n} y[i, j],$$

$$\varphi[5] = \wedge_{\substack{1 \leq i \leq k, \\ 1 \leq j[1] < j[2] \leq n}} (\neg y[i, j[1]] \vee \neg y[i, j[2]]),$$

$$\varphi[6] = \wedge_{\substack{1 \leq i \leq k, \\ 1 \leq j \leq n}} (\neg y[i, j] \vee (\vee_{1 \leq p \leq |T|-|S_j|+1} z[j, p])),$$

$$\varphi[7] = \wedge_{\substack{1 \leq i \leq k, \\ 1 \leq j \leq n, \\ 1 \leq p[1] < p[2] \leq |T|-|S_j|+1}} (\neg y[i, j] \vee \neg z[j, p[1]] \vee z[j, p[2]]),$$

$$\varphi[8] = \wedge_{\substack{1 \leq i \leq k, \\ 1 \leq j \leq n, \\ 1 \leq p \leq |T|-|S_j|+1, \\ p \leq q \leq p+|S_j|-1, \\ 1 \leq r \leq m, \\ S_j[q-p+1] \neq a_r}} (\neg y[i, j] \vee \neg z[j, p] \vee \neg x[q, r]),$$

$$\xi = \wedge_{i=1}^{8} \varphi[i].$$

It is easy to check that there is a string S such that

$$|S| = |T|,$$

$$\{S[i], S[i+1]\} = \{T[i], T[i+1]\},$$

for all $1 \leq i < |S|$, and

$$|\{i \mid \#occ(S_i, S) \geq 1\}| \geq k$$

if and only if ξ is satisfiable. It is clear that ξ is a CNF. So, ξ gives us an explicit reduction from SWCS to SAT. Now, using standard transformations (see e.g. [37]) we can obtain an explicit transformation ξ into ζ such that $\xi \Leftrightarrow \zeta$ and ζ is a 3-CNF. Clearly, ζ gives us an explicit reduction from SWCS to 3SAT.

We have designed a generator of natural instances for SWCS. We have considered our genetic algorithms OA[1] (see [38]) and OA[2] (see [39]) for SAT. For solution of SWCS, we have used heterogeneous cluster. Each test was runned on a cluster of at least 100 nodes. Selected experimental results are given in Table 1.

time	average	max	best
OA[1]	1.22 h	11.41 h	3.17 min
OA[2]	39 min	6.23 h	6.54 min

Table 1: Experimental results for SWCS.

ACKNOWLEDGEMENTS. The work was partially supported by Analytical Departmental Program “Developing the scientific potential of high school” 8.1616.2011.

References

[1] V. Yu. Popov, Computational complexity of problems related to DNA sequencing by hybridization, *Doklady Mathematics*, 72 (2005), 642-644.

[2] V. Popov, The approximate period problem for DNA alphabet, *Theoretical Computer Science*, 304 (2003), 443-447.

[3] V. Popov, The Approximate Period Problem, *IAENG International Journal of Computer Science*, 36 (2009), 268-274.

[4] V. Popov, Approximate Periods of Strings for Absolute Distances, *Applied Mathematical Sciences*, 6 (2012), 6713-6717.

[5] V. Popov, Multiple genome rearrangement by swaps and by element duplications, *Theoretical Computer Science*, 385 (2007), 115-126.

[6] V. Popov, Sorting by prefix reversals, *IAENG International Journal of Applied Mathematics*, 40 (2010), 247-250.

[7] A. Gorbenko and V. Popov, Robot Self-Awareness: Occam's Razor for Fluents, *International Journal of Mathematical Analysis*, 6 (2012), 1453-1455.

[8] A. Gorbenko and V. Popov, The Force Law Design of Artificial Physics Optimization for Robot Anticipation of Motion, *Advanced Studies in Theoretical Physics*, 6 (2012), 625-628.

[9] A. Gorbenko, V. Popov, and A. Sheka, Robot Self-Awareness: Exploration of Internal States, *Applied Mathematical Sciences*, 6 (2012), 675-688.

[10] A. Gorbenko, V. Popov, and A. Sheka, Robot Self-Awareness: Temporal Relation Based Data Mining, *Engineering Letters*, 19 (2011), 169-178.

[11] A. Gorbenko and V. Popov, Anticipation in Simple Robot Navigation and Finding Regularities, *Applied Mathematical Sciences*, 6 (2012), 6577-6581.

[12] A. Gorbenko and V. Popov, Robot Self-Awareness: Formulation of Hypotheses Based on the Discovered Regularities, *Applied Mathematical Sciences*, 6 (2012), 6583-6585.

[13] A. Gorbenko and V. Popov, Robot Self-Awareness: Usage of Co-training for Distance Functions for Sequences of Images, *Advanced Studies in Theoretical Physics*, 6 (2012), 1243-1246.

[14] A. Gorbenko and V. Popov, Robot's Actions and Automatic Generation of Distance Functions for Sequences of Images, *Advanced Studies in Theoretical Physics*, 6 (2012), 1247-1251.

[15] A. Gorbenko and V. Popov, Anticipation in Simple Robot Navigation and Learning of Effects of Robot's Actions and Changes of the Environment, *International Journal of Mathematical Analysis*, 6 (2012), 2747-2751.

[16] Z. Gotthilf, M. Lewenstein and A. Popa, On Shortest Common Superstring and Swap Permutations, *Lecture Notes in Computer Science*, 6393 (2010), 270-278.

[17] A. Gorbenko and V. Popov, On the Problem of Sensor Placement, *Advanced Studies in Theoretical Physics*, 6 (2012), 1117-1120.

[18] A. Gorbenko and V. Popov, On the Longest Common Subsequence Problem, *Applied Mathematical Sciences*, 6 (2012), 5781-5787.

[19] A. Gorbenko and V. Popov, Computational Experiments for the Problem of Selection of a Minimal Set of Visual Landmarks, *Applied Mathematical Sciences*, 6 (2012), 5775-5780.

[20] A. Gorbenko and V. Popov, The Binary Paint Shop Problem, *Applied Mathematical Sciences*, 6 (2012), 4733-4735.

[21] A. Gorbenko, M. Mornev, V. Popov, and A. Sheka, The Problem of Sensor Placement, *Advanced Studies in Theoretical Physics*, 6 (2012), 965-967.

[22] A. Gorbenko, V. Popov, and A. Sheka, Localization on Discrete Grid Graphs, *Lecture Notes in Electrical Engineering*, 107 (2012), 971-978.

[23] A. Gorbenko and V. Popov, The Problem of Selection of a Minimal Set of Visual Landmarks, *Applied Mathematical Sciences*, 6 (2012), 4729-4732.

[24] A. Gorbenko and V. Popov, The Longest Common Parameterized Subsequence Problem, *Applied Mathematical Sciences*, 6 (2012), 2851-2855.

[25] A. Gorbenko and V. Popov, Programming for Modular Reconfigurable Robots, *Programming and Computer Software*, 38 (2012), 13-23.

[26] A. Gorbenko and V. Popov, On the Problem of Placement of Visual Landmarks, *Applied Mathematical Sciences*, 6 (2012), 689-696.

[27] A. Gorbenko, M. Mornev, V. Popov, and A. Sheka, The problem of sensor placement for triangulation-based localisation, *International Journal of Automation and Control*, 5 (2011), 245-253.

[28] A. Gorbenko, M. Mornev, and V. Popov, Planning a Typical Working Day for Indoor Service Robots, *IAENG International Journal of Computer Science*, 38 (2011), 176-182.

[29] A. Gorbenko and V. Popov, SAT Solvers for the Problem of Sensor Placement, *Advanced Studies in Theoretical Physics*, 6 (2012), 1235-1238.

[30] A. Gorbenko and V. Popov, Clustering Algorithm in Mobile Ad Hoc Networks, *Advanced Studies in Theoretical Physics*, 6 (2012), 1239-1242.

[31] A. Gorbenko and V. Popov, The Problem of Finding Two Edge-Disjoint Hamiltonian Cycles, *Applied Mathematical Sciences*, 6 (2012), 6563-6566.

[32] A. Gorbenko and V. Popov, Hamiltonian Alternating Cycles with Fixed Number of Color Appearances, *Applied Mathematical Sciences*, 6 (2012), 6729-6731.

[33] A. Gorbenko and V. Popov, Footstep Planning for Humanoid Robots, *Applied Mathematical Sciences*, 6 (2012), 6567-6571.

[34] A. Gorbenko and V. Popov, Multiple Occurrences Shortest Common Superstring Problem, *Applied Mathematical Sciences*, 6 (2012), 6573-6576.

[35] A. Gorbenko and V. Popov, The Far From Most String Problem, *Applied Mathematical Sciences*, 6 (2012), 6719-6724.

[36] A. Gorbenko and V. Popov, Multi-agent Path Planning, *Applied Mathematical Sciences*, 6 (2012), 6733-6737.

[37] A. Gorbenko and V. Popov, The c-Fragment Longest Arc-Preserving Common Subsequence Problem, *IAENG International Journal of Computer Science*, 39 (2012), 231-238.

[38] A. Gorbenko and V. Popov, The set of parameterized k-covers problem, *Theoretical Computer Science*, 423 (2012), 19-24.

[39] A. Gorbenko and V. Popov, Task-resource Scheduling Problem, *International Journal of Automation and Computing*, 9 (2012), 429-441.

Received: November 1, 2012